

DOCUMENT RESUME

ED 424 766

FL 025 557

AUTHOR Salies, Tania Gastao
 TITLE Towards Communicative Measurement of Writing: Where Are We Now?
 PUB DATE 1998-00-00
 NOTE 26p.
 PUB TYPE Information Analyses (070) -- Opinion Papers (120)
 EDRS PRICE MF01/PC02 Plus Postage.
 DESCRIPTORS *Communicative Competence (Languages); *English (Second Language); Evaluation Criteria; Foreign Countries; Interrater Reliability; Measurement Techniques; Scoring; Second Language Instruction; *Test Construction; Test Items; Test Reliability; Test Validity; *Testing; *Writing Evaluation

ABSTRACT

A discussion of the evaluation of writing, particularly in English as a Second Language, argues for a communicative approach reflecting the current approach to language teaching and learning. The movement toward more communication-oriented and more valid language testing is examined briefly, and direct assessment is chosen as the preferred format within this approach. Practical procedures are then considered, focusing on possible task types, scoring, and test design. Recommended techniques include eliciting multiple samples of writing on a specific topic, holistic scoring of fulfillment of communicative intent by at least two independent raters, and realistic and concise prompts. Issues of topic selection, rater training, time constraints, test administration procedures, and test validity are also discussed. It is concluded that, while not as practical and reliable as indirect tests, such direct tests meet the goal of any language test (providing useful information about a learner's ability to effectively communicate) and exerts a positive "washback" effect on teaching and learning. A sample placement test is appended. (Contains 17 references.) (MSE)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

Towards Communicative Measurement of Writing: Where are we now?

Tânia Gastão Saliés

Pontificia Universidade Católica do Rio de Janeiro

Running head: MEASUREMENT OF WRITING

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL HAS
BEEN GRANTED BY

Gastão Saliés
Tânia

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to
improve reproduction quality.

• Points of view or opinions stated in this
document do not necessarily represent
official OERI position or policy.

FL025557

Abstract

This paper favours a communicative measurement of writing, reflecting the current approach to language teaching and learning. The opening explains this move towards communicative and more valid tests, and elects direct assessment as the ideal format within the approach. Then follows a thorough discussion of why direct assessment was chosen, a position which is supported by several research studies. Next, the paper takes a practical procedural strategy, commenting on possible methods, task types, scoring, and designing procedures. Multiple samples of writing on a specified topic, holistic scoring focusing on communication fulfillment by at least two independent raters, realistic-concise prompts are endorsed. Some final considerations on topic choice, training of readers, time constraints, and administration procedures close the body of the paper and concludes that validity should always come first. Though not quite as practical and reliable as indirect measures, direct tests meet the goal of any language test, that is, it provides useful information about a learner's ability to effectively communicate and exerts a positive "washback" effect on teaching and learning. As an addendum, a sample placement test is presented.

Towards Communicative Measurement of Writing: Where are we now?

In the not too distant past, it was our belief that language learning was synonymous with knowing grammar, structures, and endless lists of vocabulary deprived of meaning. Naturally, our testing procedures reflected that belief, and students were merely asked to reproduce memorized language, generally through recognition. Nevertheless, in the last two decades, we learned that besides linguistic rules (grammar competence), learners draw on higher order internalized systems (sociolinguistic competence, discourse competence, strategic competence; see Canale and Swain 1980; Bachman 1991) to effectively communicate. These different systems combine to structure and give social and propositional meaning to language, and meaning becomes a function of the interaction among the linguistic code, functions, and context, entailing the intentions of the speaker and the expectations of the hearer. In other words, language is a whole, not the sum of discrete syntactic, phonological, morphological, semantic, discourse, and organizational parts. Consequently, if we aim at measuring the writing ability of our students communicatively, we must test all the mentioned levels of competence, triggering the examinees' grammar of expectancy, reflecting real-life language use, designing tasks that require more than simple knowledge recognition, or a "yes" or "no" answer, but the actual performance of the trait relative to the objectives of the test itself, and the needs of the learner.

In brief, it is my understanding that a communicative and valid measure of writing tests *production*, not knowledge recognition; activates the internalized rule

systems simultaneously, not discretely; meets specific language needs in a given circumstance, as defined in the objectives of the test; manipulates a variety of language functions; stresses communication and meaning; and uses format and scoring procedures that reflect this understanding, drive curriculum progressively, and create the conditions under which good writing is known or is apt to occur. Ideally, then, our tests should be direct, or performance-based.

A rationale for direct measures of writing

In consonance with the communicative paradigm, the majority of research studies and language teachers today widely support direct measures of writing, among other things, for its validity, authenticity, and instructional role.

To begin with, literature is conclusive about the importance of validity and its primacy over reliability. For instance, Quellmaz (1982), Cooper (1984), Brossell (1986), Stansfield & Ross (1988), Ruth & Murphy (1988), Greenberg (1990), and Hughey (1990) note that we should first require our students to develop content, organize ideas, use appropriate vocabulary and syntax, drawing on their higher order systems to convey meaning, and then attempt to make these measures as reliable as possible, limiting possible sources of error as task type, topic selection, timing, and scoring procedures. These research studies are positive about the importance of measuring the right "thing," even if with some inconsistency. They support the contentions that direct measures tap a production factor, and thus represent a separate construct from that of indirect tests (namely, the ability to write as opposed to knowledge of conventions of writing). Indeed, nothing seems

more logical than requiring students to actually write to gauge if they can do it. If we want to find out if young people can swim, we simply ask them to jump into a pool and swim. Why don't we do it with writing?

Furthermore, some of these research studies -- Brossell (1986), Cooper (1984), Ruth and Murphy (1988), Greenberg (1990), and Hughey (1990) -- further supported by Lutz (1983) and Wesche (1987), indicate in their rationale the meaningfulness of direct assessment. They contend that it reproduces real-life communication acts, using other participants, the scorers, to judge the success or failure of the writer's communicative efforts. Simply stated, they argue for a direct measure of writing because it is authentic. It gives social and propositional meaning to language. It demands negotiation of meaning, and awareness of the reader.

Finally, Cooper (1984), Brossell (1986), Wesche (1987), and Greenberg (1990) bring to light a third reason for using a direct measure of writing: its instructional role. Tests directly influence what is taught, and consequently, what is learned. Therefore, we should use them as tools to provide growth in knowledge, and greater skill in writing, progressively driving the curriculum. If teaching to the test occurs, it is far more desirable to have teachers training students to pass a writing sample than an objective test. After all, students may end up learning to write by simply trying to write. If for nothing else, this is a sufficient reason to adopt direct writing tests.

In conclusion, although this review is very modest and obviously full of gaps, it leaves no doubt about the validity, authenticity, and instructional importance of direct measures within a communicative framework. It presents, I believe, substantial support for my standpoint, although divergencies relative to number of samples, format, reliability, and task/topic, cost issues remain to be equated.

Types of direct measurement

This section intends to be a brief summary of the various types of direct measurement available. Test users should choose those most appropriate and authentic within course objectives and needs of the students, taking into account, communicatively speaking, that we must test what the examinee will actually have to do in a naturalistic situation. To put it simply, if I am testing academic English of ESL graduate students, it is not realistic to ask them to write a personal letter, but to argue and take a position on a general topic, a task they will have to perform constantly in the academic environment.

The types of direct assessment commonly used may be classified according to methods of elicitation and task types. Among the *methods of elicitation*, the essay test is the most common and traditional method for getting students to write (Weir, 1990). Topics are often general, easy to understand, personally related, and not biased towards any specific group or content area. No clues on how to answer the question are provided. Secondly, there are controlled writing tasks. This method avoids the variety of approaches candidates tend to have towards open-ended stimuli, specifying media, audience, purpose, and situation through written,

spoken or non-verbal stimuli (a graph, for example, as administrations of the TWE used to do in the late 80's). If the task is determined, it is easier to compare performances of different students, and obtain higher reliability in scoring. Nevertheless, in some cases, if we determine the task, we restrict creativity and draw on other skills (prompt interpretation, ability to understand graphs or charts, for instance), sacrificing validity somewhat. In the case of the TWE, it is designed to test graduate and undergraduate students of different academic backgrounds. Therefore, the graph prompt proved extremely inadequate for incoming English undergraduate students, since it draws on the ability to understand histograms, pie charts, or statistical data that some of these students might have never dealt with. It ended up being discontinued by ETS. Finally, a real-life task of some importance is that of synthesizing information (mainly in the academic environment): the summary test method (Breland, 1983; Weir, 1990). It involves the ability to write a controlled composition that contains essential ideas and omits non-essentials, through re-combination of data in an acceptable form. Indeed, it is a crucial important skill for students in an academic situation, but it presents several difficulties as selecting an appropriate-unbiased-general passage; scoring reliability (even with an answer key with the main points of the passage, some subjectivity still remains); and, depending on the population, suspect validity (adults who use the language for everyday purposes don't need to develop this academic skill).

Task types, on their part, vary with topics and prompts used to elicit the desired language behaviour (modes of discourse). Among some well-known types of writing tasks, I would cite narratives (real or imaginary, it could be an autobiographical account, a description of some sort, etc.); descriptions (it implies description of a series of events, of an object, how it looks or works); argumentations (the most common in essay tests, because it asks examinees to take a position on some issue and to argue persuasively using their own personal experience, integrating different writing skills); and expositions (expository in nature, but it only requires an opinion on some issue or event). The TWE (Test of Written English by ETS), for example, after extensive survey about the field-specific writing demands in American universities, uses either the compare-contrast and take-a-position task (argumentative essay). The MELAB (Michigan English Language Assessment Battery) contains a writing test which consists of either a personal narrative or of an argumentative-take-a-position task.

Ideally, to provide a fairly representative sample of the examinees' writing ability, a writing assessment should present at least two prompts, independent from one another (Godshalk, 1966; Wesche, 1983; Quellmaz, 1982; Breland, 1983; Cooper, 1984; Pollitt & Hutchinson, 1987; Stansfield & Ross, 1988; Greenberg, 1990; and others). Some examinees are likely to perform better at some tasks than at others. In doing so, we control those contextual features that determine difficulty, cover a broader range of language functions as defined by Finocchiaro & Brumfit (1983), and enhance validity and reliability. For instance, TELS --The

English Language Skills Profile (Hutchinson & Pollitt, 1983) -- uses five different tasks: writing a letter, writing a report, writing a newspaper article, imaginative story telling, and expressing an opinion. ELTS -- English Language Testing System -- uses two: describing a diagram/graph/drawing, and writing a report/argumentation on the passage of the reading section of the battery. Of course, such models are expensive and time-consuming (doubtless, excellent models for their purposes/needs), and may prove impractical for large-scale testing. In this case, as Greenberg (1990) in her analysis of the TWE mentioned, it is better to have one writing sample than none, emphasizing the importance of positive backwash, and construct validity.

Scoring procedures

There is much disagreement on the approaches and descriptions of writing evaluation methods. Based on studies conducted by Jacobs *et al.* (1981), Weir (1990), Hughes (1989), and Breland (1983), I will describe two basic scoring processes: holistic and analytic, favouring the first for its communicative approach, and practicality.

In *holistic scoring*, markers base their judgments on the impression of the whole composition. Cooper (1984) defines it as any procedure which stops short of enumerating linguistic, rhetorical, or informational features of a piece of writing. This means, not focusing on mechanical or grammatical weaknesses of the writing sample, but on its overall impression; attending to the writer's message; staying closer to what is essential in realistic communication. For example, one might

score for content, organization, and language usage without specifically focusing on any of these aspects in particular, but on the final result produced by their combination in the effort to successfully convey meaning. It is essential to observe that the subjectivity of marking must be controlled to strike a balance between reliability and validity. Some necessary steps in this direction are: the establishment of defined criteria for each level of performance; double scoring (at least); and previous training of raters. Although such subjectivity in reading essays was long thought undesirable, it has become a strength within a communicative approach, because it entails meaning negotiation, and is part of any communication act. In spite of it, holistic scoring has shown high reliability results. Jacobs *et al.* (1981) indicates that most research studies found it to be in the mid-to-high eighties or nineties when raters are well-trained on the established criteria.

In *analytic scoring*, on the other hand, the focus is on distinct aspects of language, as for example, content, organization, language usage, mechanics, etc. Each aspect is scored separately, and then summed up in a total score. Because the rating criteria is usually more explicitly defined, it is a more objective and reliable method. Nevertheless, of suspect validity (Weir, 1990), because it evaluates parts, not the complete picture of the learners' performance (communicative effectiveness); and it is less economical (more time-consuming).

I will observe, however, that several considerations, other than framework, must be taken into account when choosing the evaluation method -- purpose of the test, accuracy required, practical constraints (time, money, personnel availability),

and type of task, to name just a few. Hughes (1989), and Weir (1990), note that if the purpose is to rank students (placement tests), direct assessments with holistic scoring are clearly valid measures; but if the purpose is to identify strengths and weaknesses of a student's writing for instructional feedback (diagnostic tests), analytic scoring coupled with an additional impressionistic score are required, if we intend to be coherent with our framework. Since this may prove economically impossible, it is my belief that the best criteria to use is a holistic scoring guide which assigns a single score for the communicative effect of combined writing skills at each specific level. The scale and its descriptors ought to be established according to the objectives of the test. The British Council's ELTS test, for example, presents nine bands with accurate descriptors, because it was designed to assess if a student's writing ability is adequate for study in English in a British university; in another instance, The American Council for the Teaching of Foreign Languages, ACTFL test, measures against six bands with highly detailed descriptors, because it aimed at providing additional criteria to assess foreign language learning in schools and colleges; and the TWE uses 6 bands with general descriptors, because it focuses on the writer's overall writing competency.

Preparing the writing task

Considerations of practicality, reliability, validity, and test purposes set the parameters for designing the writing task. Basically, in order to yield reliable and valid results, and yet be within practical constraints, the task should be realistic, appropriate, understandable, personal, feasible, representative, and fair. In other

words, the task should reproduce a real-life composing situation in terms of knowledge and discourse mode, involving the writers, giving them a chance to write on a subject they know and are interested in. It should be compatible with writers' educational level, cultural and socio-economic backgrounds. It should be briefly, objectively, and simply stated (we are testing writing, not reading. People decode messages in the most unexpected ways!) It should be motivating in the sense that it triggers the writer's own perception of the topic. It should be workable for both the writer and the reader within the amount of time assigned for it. It should provide an adequate sample of the writer's ability, preferably providing more than one opportunity to write (fresh-starts) through different modes of discourse, thus neutralizing difficulty and psychological factors from topic to topic, or from one test period to another. Finally, it should not be biased towards a specific content area, or cultural group.

Further considerations

There are a number of other factors that influence the performance in a direct writing test, introducing variance into it, and affecting its validity and reliability. For example, *topic choice*. In order to ensure comparability among students, and thus, enhance reliability, it is generally advisable to have all of them write on the same topic (Jacobs *et al.*, 1981; Godshalk *et al.*, 1966; Brossell, 1986; Quellmaz, 1982; Cooper, 1984). Otherwise we may be favoring some students in different respects: Different subjects demand different vocabulary, knowledge, organization structure and tone. Secondly, *training and number of essay readers*. At least two

experienced readers, trained on the criteria established, and on the scope of the prompt topic, should read the composition rapidly. Readers themselves may interpret the task in different ways. So it is important to have a consensus on how and what to be looking for. Preferably, they should be ESL English teachers, with experience in grading compositions. Thirdly, *time constraints*. Ideally there should be no time limit to let writers demonstrate their abilities to the most, reproducing a life situation. Nevertheless, it is not a feasible solution. The number of writing samples, and the size of the test group will influence the decision about the amount of time. Normally, large-scale tests (like the TWE and the MELAB) assign 30 minutes for one single prompt. The ELTS assigns 45 minutes for two tasks (15 for task number one, writing a description of a graph/chart/drawing, and 30 for task number two, writing a report/argumentation). When testing smaller populations, ESL teachers and researchers have reported (Jacobs *et al.*, 1981) a range of ten minutes per task, for a total of four short essays, and up to thirty minutes per task in the case of college students. Fourth, *administration procedures*. Every administration must provide fair and equivalent conditions to avoid the introduction of systematic errors in score variance. For example, time of the day, day of the week, conditions of heating and lighting, persons monitoring the exam, and so on.

Conclusion

Writing is too complex a skill to be measured through discrete point tests. It involves so many sub-skills and cognitive processes that an integrative and direct

test is a demand. Unless we want to measure the wrong "thing," sacrificing construct validity. Besides, if we want this measure to be in tune with the communicative paradigm, students must be required to negotiate meaning, exercising organizational, pragmatic, and strategic competencies in the actual performance the trait, as they would in real-life situations.

Controversial and pervasive issues such as reliability, cost, and time should no more intrude in the decision for direct assessment. Quality should always come first. Validity, after all, is essential. And besides, subjectivity is a natural and unique characteristic of any communication act. Why not of writing tests which intend to measure communicative effectiveness?

We should, therefore, be attempting to obtain as many samples of our students' writing as permitted by practical constraints, keeping in mind that a single and brief communicative sub-test is better than none (if for nothing else, for its positive effect over teaching and learning); that reliability can always be enhanced by careful selection of tasks; focused holistic evaluation criteria; previous training of readers; multiple ratings, a wide enough sample of language functions; and even, in the case of large scale tests, through a combined format -- multiple-choice sections, followed by an essay (as in the MELAB, and the TWE).

In short, nothing can substitute for the practice of writing. And we will only acknowledge its importance in the curriculum and encourage its cultivation by adopting direct assessments as our testing "modus operandi."

References

- Bachman, Lyle. (1991). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Breland, H.M. (1983). The direct assessment of writing skill: A measurement review. (ERIC Document Reproduction Service no. ED 242 756)
- Brossel, G. (1986). Essay test topic development (ERIC Document Reproduction Service no. ED 279 002)
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1 (1), 1-47.
- Cooper, L.P. (1984). The assessment of writing ability: A review of research. (ERIC Document Reproduction Service no. ED 250 332)
- Finocchiaro, M., & Brumfit C. (1983). *The functional-notional approach: From theory to practice*. New York: Oxford University Press
- Godshalk, F.I., Swineford, F., & Coffman W.E. (1966). The measurement of writing ability. *ETS Research Monograph*, 6. Princeton College Entrance Examination Board.
- Greenberg, K. (1986). The development and validation of the TOEFL writing test: A discussion of TOEFL research reports 15 and 19. *TESOL Quarterly*, 20, 531-44.
- Hughey, J. (1990). ESL composition testing. In *English Language Testing in U.S. Colleges and Universities*, 6, pp. 51-67. (ERIC Reproduction Document Service no. ED. 331 284)

- Hughes, A. (1989). *Testing for language teachers* (pp. 75-100). Cambridge: Cambridge University Press.
- Jacobs, H.L., Zingraf A.S., Wormuth R.D., Hartfield F.V., & Hughey B.J. *Testing ESL composition: A practical approach*. Rowley, Mass: Newbury.
- Lutz, W. (1983). What we know and don't know: Needed research in writing assessment. (ERIC Document Reproduction Service no. 236 242)
- Pollitt, A., & Hutchinson, C. (1987). Calibrating graded assessment: Rasch partial credit analysis of performance of writing. *Language Testing*, 3 (1), 72-92.
- Quellmaz, E.S. (1982). Designing writing assessments: Balancing fairness, utility and cost. (ERIC Document Reproduction Service no. ED 228 270)
- Stansfield, W.C., & Ross, J. (1988). A long-term research agenda for the Test of Written English. *Language Testing*, 5 (2), 160-86.
- Weir, J.C. (1990). *Communicative language testing* (pp.58-73). New York: Prentice Hall.
- Wesche, B.M. (1987). Communicative testing in second language. In M.Long & J. Richards (Eds.), *Methodology in Tesol: A book of readings* (pp. 373-394). Rowley, Mass: Newbury.

Addendum

Having the discussed considerations in mind, I designed a placement writing test for EFL would-be English teachers entering the academic life in Brazil. A high level mastery of grammatical, discourse and sociolinguistic components are required from them. Its purpose is to assign these incoming undergraduate students (who were pre-approved in a general proficiency admission test to the Teachers Training Course of the English program) to composition classes at three levels: Composition I (basic composition skills on paragraph development, different modes of discourse, and organization patterns, idioms/ usage); composition II (review of basic composition skills onto essay writing); Composition III (instruction in research/technical writing, usage/grammar review). The test is syllabus based, and writers' success in communicating clearly their ideas through a well-organized composition, with language appropriate to the task, and with good control of language mechanics is given greater attention. Sufficient accuracy is required to avoid too many changes once classes are underway. Backwash is a serious consideration. Around 20 students are admitted to the program each semester (January / July); therefore, time and scorers availability (EFL composition faculty, during summer and winter vacation) is a medium constraint. The test assesses at least two functions of language, providing a broad enough sample of individual students' performance.

*Test Specifications*1. *CONTENT*

Methods of elicitation, task types, and topics should reflect the kinds of written texts found in the institution's English program where English is a foreign language, being as neutral as possible. The student is expected to write to native, and non-native EFL composition faculty, as well as to the English speaking community in general. As we are interested in measuring performance (vide course/test objectives), testing is direct, and include several levels of cognitive processes, and underlying skills. Namely, drawing on their knowledge of the world to organise and present information; describing a picture/events/objects or persons; narrating events; exposing ideas, persuading and taking a position; developing a thesis, topic sentences, adequate support, and transitions; using the conventions of the language (spelling, and punctuation), idioms, sentence construction, word order, verb agreement, prepositions, articles, and appropriate vocabulary.

2. *FORMAT AND TIMING*

Students are asked to complete two essay tasks. One involves information reprocessing (15 minutes). They might be asked to look at a diagram, a drawing, or a piece of text and to present the information in their own words, in a coherent and cohesive piece of writing. The other requires them to draw on their own experience and knowledge of the world to expose/argue/report/narrate on a topic (45 minutes). They do not have a choice

of topic, and they must do the two task types presented. The text types are purposefully broad in order to encompass the course syllabus, and exert a positive effect on learning if practice for the test occurs. Besides, they are encouraged to plan and organize their writing in the first minutes of each task. Total testing time: 1 ½ hour. Test topics are printed on separate pages, with complete and clear instructions. Ruled paper is provided to make writing and reading easier, thereby facilitating scoring and enhancing reliability. There are no specifications relative to length, but in general, students are expected to write one front page on the first assignment, and one full page (front and back) on the second. The test is administered in the morning, and students are identified by their ID number. Below, samples of the two task types (please note that a prompt cannot be used in more than one administration).

Samples of task one:

1) Pictures normally have an effect on people. Describe the one you see below. Build your description around a particular feeling or tone to let the reader know what your impression of it is.



2) The chart below shows some people's commonly observed behaviors. Using the information it provides, **compare** the value people at different stages of life place on different behaviors and **take a standpoint**.

PEOPLE	BEHAVIORS
Children	To enjoy life
Adolescents	To complain
Young-adults	To make money
Middle-aged	To live family life
Aged adults	To enjoy life

3) Look carefully at the sequencing of pictures below. Write a small story about what they tell you.

A



B



C



D



(Byrne 1967)

Samples of task two:

- 1) Preparing for end of year examinations involves both long-range and short-range planning. Using one or two examples, compare the two ways. Which way is your favorite? Why?
- 2) "Words alone do not make a language." What kind of arguments could you use to support or refute this point of view? In a well-developed essay, discuss your position.
- 3) There is no denying that English is a useful language. Write a well developed essay on the multiple uses of knowing it nowadays. Give at least three examples.
- 4) Ecologists' predictions of a major ecological disaster do not seem far-fetched if you consider the world's population who are starving. Write a well-developed essay on the steps you consider important to move towards a more ecologically responsible world. Give at least three examples.

3. SCORING

Writing samples will be scored by EFL composition faculty, trained in the test procedures, through a holistic scoring guide. They will use compositions written by enrolled students to practice and obtain an inter-rater reliability coefficient of .90. Each composition is read twice, quickly (three minutes for each), by two independent raters. The first time, to form an overall impression of the communicative effectiveness of the piece of writing. The second, to ascertain that the criteria established by the guide were correctly applied. If the raters

disagree, a third rater will be called in. The holistic scoring scale is broken down into three mastery levels in consonance with the expected performance in the three composition courses offered.

SCORING GUIDE

ID #	Date	Task 1	2
COMPOSITION III <i>Demonstrates competency in writing, addressing the task effectively. Fluent expression; well organized (thesis/topic sentences); thorough development of the topic with adequate support, concise and effective wording, and logical sequencing. There may be occasional mechanic errors, but there is full command of the conventions of the English language.</i>			
COMPOSITION II <i>Demonstrates some competency in writing, but addresses the task partially. Main ideas stand out, but with limited fluency: topic not fully developed; desiring organization (thesis/topic sentences not very clear); logical but incomplete sequencing of ideas; and inadequate support (lacks detail). Occasional mechanic and word/idiom errors don't obscure meaning.</i>			
COMPOSITION I <i>Demonstrates little competency in writing. Non-fluent expression; ideas not clearly stated; inadequate topic sentences and development of the topic; inappropriate/insufficient details or logical wording; frequent mechanic and word/idiom errors obscure meaning.</i>			
Score	I	II	III

The examinee is expected to obtain the same level of performance on both tasks to be placed at that level. If not, the lower level prevails.

4 ***SAMPLING***

Task types are supposed to represent a wide sample of the specifications under content. Succeeding versions of the test should do the same.

5 ***ITEM WRITING AND MODERATION***

Writing tasks should be set through teamwork. EFL faculty should work together to validate them, trying hard to find fault. Critical questions as is the task specific enough? Is it clear, concise, brief? Is it testing anything else besides writing skills? Is the topic neutral enough? Is it eliciting the behaviours it intends to measure? Does it reflect the course syllabus?

6 ***PRETESTING***

Several tasks will be designed. All of them will be pretested on current students enrolled in the three composition levels to check for problems in design, administration and scoring. Each score will be compared with the student current level in the program. If there should be a problem in the critical levels, or with task/topic selection, improvements should be done during this phase. Specific items will only be re-used one year after they had been used in a pretest.

7 ***VALIDATION***

The test will be validated against the proportion of students placed inappropriately (criterion-related validity).

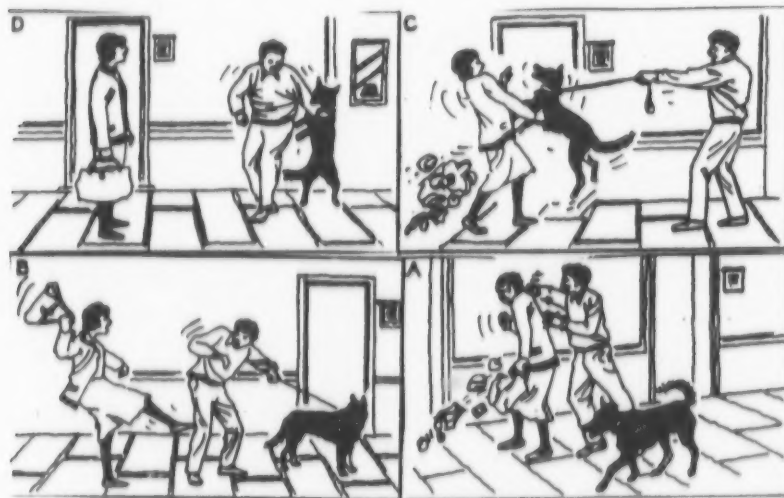
8 ***SCORES MODERATION***

Inter-rater reliability will be computed. It should be a strong, positive, and significant correlation (ideally around .90).

SAMPLE WRITING TEST

This is a test of your ability to write in English. Take some time to plan and organize your ideas. There are two tasks, both of which must be completed. You will have 30 minutes to spend on task 1, and 60 minutes to spend on task 2. Make sure you skip every other line.

1. Look carefully at the sequencing of pictures below. Write a thirty minute story on what they tell you.



This image shows a single sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There is no handwriting or other markings on the paper.